Disordered Heteropolymers with Crosslinks - Phase Diagram and Conformational Transitions

Lorin Gutman and Eugene Shakhnovich

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

We study the phase behavior of random heteropolymers (RHPs) with quenched cross-links, a novel polymer class of technological and biological relevance, and show the possible occurrence of freezing with few chain conformations sampled. The sensitivity of the frozen phase microstructure to the disorder components is elucidated at positive solubility parameter values; at low T's segregated microphases form, while at a finite T, a first order conformational transition occurs, and is attributed to statistical matching of large microphases bounded by cross-links. The end of the symmetry broken regime stabilization by cross-links occurs at a higher T by a second order conformational transition. (*submitted to Chem. Phys. Lett.*)

Random Heteropolymers (RHP) with some monomers bonded by quenched cross-links may be viewed as a natural offspring of two structurally remote classes of polymers, cross-linked homopolymeric networks [1]− [3], and linear RHPs [4]− [6], both renown for their impact on technology and biology. The theoretical analysis of cross-linked RHPs is eminently challenging; two frozen disorders are carried by the polymer; the first disorder component is the sequence distribution of disparate monomers that has been fixed during synthesis: for linear RHPs known manifestations of having a fixed sequence of segments in polymer properties are mesophase formation due to spatial grouping of monomers with similar composition [4]− [5], and the reduction of overall chain conformations to a small number of predominant folds (frozen phase formation) [6]− [9]. Analytical [4]− [9] and computer simulations [10]− [13] studies of linear RHP chains clarified major questions on protein folding possible mechanisms, thermodynamics and provided insight into the dynamic pathways of folding of proteins to their native conformation. The main past study emphasis on crosslinked chains has been on crosslinked homopolymers, the earliest proeminent work in the field has been by Deam and Edwards [1]; recent advances attained in the understanding of cross-linked homopolymer networks have been discussed and reviewed by Goldbart and co-workers [3] and Panyukov [2].

A detailed modeling of crosslinked RHPs is challenging, while complex phase behavior is expected in comparison with linear RHPs; the quenched sequence distribution of segments fixed by the chain connectivity of the incipiently linear RHP was shown [6] to preclude macrophase separation by formation of microdomains. Crosslink formation in linear RHPs, and the nature of these crosslinks (quenched or annealed) are expected to significantly impact the formation and stability of microphases, and their interdomain interfaces as well as the overall number of possible accessible chain folds.

The weaker disorder - the sequence distribution of segments reminiscent of linear RHPs, and the strong disorder component - the fixed junction network reminiscent of crosslinked homopolymers, both present in our system <u>cannot</u> be treated as annealed disorders. This important issue is addressed in the present work by the extension of spin glass methodology of analysis of systems with one quenched disorder [14]; our formalism allows the explicit analytic treatment of the quenched nature of both the sequence and crosslink distribution. The phase behavior, the conformation chain organization in these phases, the interrelation of spatial monomer separation by composition to the occurrence of phases of few dominant folds, and the transition between these phases is studied.

Theory Development

Imagine a dilute solution of statistical two-letter heteropolymers. Composition specific crosslinking agents are introduced in solution and intrachain composition specific junctions form; the A and B segments residing in proximal spatial range can form homogeneous crosslinks (A-A and B-B type) and heterogeneous ones (A-B type). Upon completion of the crosslink synthesis, the crosslinks are quenched by chemical means and no further crosslink reorganization can occur. For example in proteins crosslinks are quenched by irreversible reactions with iodoacetamide or by acidification [15].

The microscopic interactions for the incipient two-letter linear RHP are described by a continuous microscopic Hamiltonian representation inspired from Edwards work on homopolymers [16]. $\mathbf{r}(n)$ is the spatial location of the n'th segment, while $\theta(n)$ monitors the chemical composition of the n'th segment. $\theta(n) = 1$ for an A segment and $\theta(n) = -1$ for a B segment. The coarse-grained description of the segment composition fluctuations for statistical RHPs along the chain contour obeys a Gaussian process with mean, $< \theta >= 2f - 1$, and sequence fluctuations [17]− [18], $< (\delta\theta(n)\delta\theta(n')) >= \delta(n - n')4f(1 - f)l$; l is the statistical segment length, and f is the fraction of A segments. The Hamiltonian is:

$$H'_{RHP} = \frac{3}{2l} \int \dot{\mathbf{r}}^2 dn + \frac{1}{2} \sum_{i,j=A,B} \int d\mathbf{r} \int d\mathbf{r}' \hat{\rho}_i(\mathbf{r})\mathbf{V}_{ij}\hat{\rho}_j(\mathbf{r}) \quad (1)$$

The first term in eq. 1 represents the nearest neighbor interaction and accounts for the chain flexibility. In the second term of eq. 1, $\hat{\rho}_i(\mathbf{r})$ represents the microscopic density composition components given by:

$$\hat{\rho}_A(\mathbf{r}) = \int dn \frac{1}{2}\left(1 + \theta(n)\right)\delta(\mathbf{r} - \mathbf{r}(n)) \;\; ;$$
$$\hat{\rho}_B(\mathbf{r}) = \int dn \frac{1}{2}\left(1 - \theta(n)\right)\delta(\mathbf{r} - \mathbf{r}(n)) \qquad (2)$$

$\mathbf{V}_{ij}$, the binary inter-segment interactions, are represented by the following matrix:

$$\overline{\mathbf{V}}_{ij} = \begin{pmatrix} V_{AA}(\mathbf{r} - \mathbf{r}') & V_{AB}(\mathbf{r} - \mathbf{r}') \\ V_{BA}(\mathbf{r} - \mathbf{r}') & V_{BB}(\mathbf{r} - \mathbf{r}') \end{pmatrix} \qquad (3)$$

$V_{AA}(\mathbf{r}\text{-}\mathbf{r}')$, $V_{AB}(\mathbf{r}\text{-}\mathbf{r}')$ and $V_{BB}(\mathbf{r}\text{-}\mathbf{r}')$ represent A-A, A-B, and B-B segment-segment interactions, respectively. Quenched composition specific crosslinks of type A-A, B-B, or A-B are described in our theory as instantaneous spatial constraints imposed on the partition function of a linear RHP. A general composition specific crosslink has the form $\delta(\mathbf{r}_i(n) - \mathbf{r}_j(n'))$ where i,j =A,B.

Fluctuations in the total number of cross-links between different disorder realizations of crosslinks are allowed by a Poisson distribution; this choice has been recently discussed [19], [20], [21] in context of homopolymer network studies. Thus, a fixed number of A-A, A-B and B-B cross-links is linearly parameterized by μ_{AA}, μ_{BB} and μ_{AB} , and the probability of fluctuations around μ_{AA}, μ_{BB} and μ_{AB} is given by:

$$P[\mu's] = \frac{[\mu_{AA}]^M}{M!}\frac{[\mu_{AB}]^J}{J!}\frac{[\mu_{BB}]^K}{K!}$$
$$exp[-(\mu_{AA} + \mu_{BB} + \mu_{AB})] \qquad (4)$$

Thus, the partition function of an RHP chain with a fixed sequence and constrained by M crosslinks of type A-A, K crosslinks of type B-B, and J cross-links of type A-B is given by:

$$Z[\theta(n)] = \underline{\int\int}D\mathbf{r}(n)exp(-\frac{3}{2l}\int\dot{\mathbf{r}}(n)^2 dn +$$
$$\frac{1}{2}\sum_{i,j=A,B}\int d\mathbf{r}\int d\mathbf{r}'\hat{\rho}_i(\mathbf{r})\mathbf{V}_{ij}(\mathbf{r} - \mathbf{r}')\hat{\rho}_j(\mathbf{r}))$$
$$\left[\frac{1}{2^M M!}(\int d\mathbf{r}(\hat{\rho}_A(\mathbf{r})^2)^M\right]\left[\frac{1}{2^K K!}(\int d\mathbf{r}\hat{\rho}_B(\mathbf{r})^2)^K\right]$$
$$\left[\frac{1}{J!}(\int d\mathbf{r}\hat{\rho}_A(\mathbf{r})\hat{\rho}_B(\mathbf{r}))^J\right] \qquad (5)$$

In eq. 5 we have expressed the crosslinking constraints by microscopic sequence-dependent-composition-densities (viz. eq. 2). The spatial $\mathbf{r}$ integration and the coefficients in front of the constrains accounts for formation of all possible crosslinking constraints that are consistent

with crosslinking from an equilibrium ensemble of composition specific contacts, and reminiscent of the typical experimental chemical crosslinking process considered here.

The free energy computation of our problem requires to average the logarithm of the partition function, Z given in eq. 5, since both the sequence and the crosslinks are quenched:

$$F = \sum_{[\theta],[r_i]} P_1([\theta])P_2([\theta],[r_i])log(Z([\theta],[r_i])) \qquad (6)$$

$[\theta]$ represents one fixed sequence realization while $[r_i]$ are the spatial coordinates of one fixed crosslink realization. $P_1([\theta])$ is the probability distribution for the synthesis of one quenched sequence, while $P_2([\theta],[r_i])$ is the conditional probability distribution of one specific realization of composition specific cross-links given a preexisting fixed sequence; thus at each given sequence realization, crosslinks can form by fixing some composition specific intersegment contacts from spontaneously occuring chain conformations adopted by the linear RHPs.

The crosslink average is performed first by introducing 3(n+1) identical copies of the system; this mathematical trick originally introduced by Deam and Edwards [1] in their studies of crosslinked homopolymers, is an exact and non-perturbative way for fixing intersegment contacts from spontaneously occuring binary contacts from an equilibrium distribution of chain conformations in homopolymers [1]:

$$F([\theta]) = \sum_{[r_i]} P_2([r_i],[\theta])log(Z([\theta],[r_i]) =$$
$$\left(\frac{\partial}{\partial n}\right)_{n\to 0} log < (Z^{n+1}) >_{cl} \qquad (7)$$

Our definition of this averaging trick is slightly different from Deam and Edwards formulation; in our formulation the derivative of log(Z) allows in the present problem an exact representation of the sequence dependent denominator occuring due to the probability normalization, in a nominator form, and sets the ground for an exact analytical sequence average later on.

The crosslink average of the log of the partition function given in eq. 5 is performed with the Poisson distribution given in eq. 4 by summation over M, J, L values.

Next we perform the average over disordered sequence; the log generated in the crosslink average (viz. eq. 7) is averaged now over the sequence by using one more time our averaging trick; thus, the free energy is now given by:

$$F = \left(\frac{1}{m}\right)_{m\to 0}\left(\frac{\partial}{\partial n}\right)_{n\to 0} << (Z^{(n+1)\otimes m}) >>_{sq,cl} \qquad (8)$$

Interestingly, the sequence average of the log generates in turn a new index m of copies.

All together, $m \otimes (n+1) = l$; l is the total number of copies of our system. Methods developed in the field of spin glasses by Parisi [23] are extended herein, and used to obtain an analytical solution. The sequence distribution average is carried out here exactly in the usual way [17].

As in previous RHP studies, cross-linked RHPs in compact states are of interest; under these conditions a scaling argument shows [22] that it is reasonable to compute the free energy by a one-step mean-field Parisi [23] calculation. It was shown that the reduction in the total number of chain conformations to a small number of dominant folds, a phenomena usually termed in the RHP literature as freezing into a few chain conformations, is described by one order parameter ξ; the polymer coordinates, (the annealed degrees of freedom) equilibrate in response to the quenched disorder realizations regardless of the disorder source, sequence, crosslinks or both; thus, in the present case one freezing order parameter ξ_0 measures the total reduction of chain conformations due to energetic and entropic constraints. The physical meaning of ξ is as follows:

$$\xi_0 = 1 - \sum p_i^2 \qquad (9)$$

p_i is the probability of the i'th chain conformation. For many chain conformations, each conformation has a low probability of occurrence, and ξ_0 practically equals one. If the chain is collapsed in one conformation, p_i=1, while for other chain conformations where i$\neq$ j, p_j=0 implying that ξ_0=0.

In the present case, not like in the linear RHP case, ξ is an explicit function of the indices, (α, k). A suitable one step parameterization of the freezing order parameter ξ, is $\xi_0 = x_0 x_0'$; m and n form a tensor space, $l = m \otimes n$; x_0 and x_0' are the one step continuous parameterizations [23] of the indices α (crosslinks sector), and k (sequence sector), respectively. A separate publication is now in print wherein all mathematical details and derivations are presented [22]. The scenario of $x_0' = 1$, $0 < x_0 < 1$ is defined as sequence-induced-freezing, the scenario $x_0 = 1$, $0 < x_0' < 1$ is defined as crosslink-induced-freezing while the parameter regime of $0 < x_0 < 1$, $0 < x_0' < 1$ is defined as sq.+cl.-induced-freezing as obviously inferred by the parameterization relevant in determination of the overall number of folds. These convenient definitions allows to keep a good track of the numerical results while we discuss the physics but should not be taken too literally; first, since the true order parameter that monitors the occurrence of few dominant folds is ξ_0 while x_0 and x_0' are parameterizations only; also, as our calculations show, and is inferred by the crosslinking procedure the disorder components are strongly correlated, and it is not possible to completely separate the component disorder manifestations in the conformational organization of the crosslinked RHP.

Using Parisi commutation relations [23], the total free energy per monomer for an RHP in compact state with composition specific and quenched cross-links is:

$$F = \frac{\gamma}{x_0 x_0'}(1 - 2(x_0 + x_0') +$$
$$(2x_0' - 1)(x_0 - 1)(x_0 - 2)^2)$$
$$+ \frac{log(1 - \mu_a x_0 x_0')}{x_0 x_0'} - \mathcal{V}(\mu's, f)(x_0 - 1)$$
$$\mu_a = \frac{\sigma^2}{2}\rho\overline{\chi(\mu's)}; \gamma = \frac{l}{0.25(2V_{AB} + V_{AA} + V_{BB})} \qquad (10)$$

with:

$$\mathcal{V}(\mu's) = 0.5(\mu_{AA} + \mu_{BB} + 2\mu_{AB})$$
$$\mathcal{V}(f, \mu's) =$$
$$0.5\mathcal{V}(\mu's) - 0.25(\mu_{AA} + \mu_{BB})\overline{\theta} - 0.25\chi_{Fcl}(\mu's)\overline{\theta}^2$$
$$\chi_F = 0.5(2V_{AB} - V_{AA} - V_{BB})$$
$$\chi_{Fcl}(\mu's) = 0.5(2\mu_{AB} - \mu_{AA} - \mu_{BB})$$
$$; \overline{\chi(\mu's)} = \chi_F - \chi_{Fcl}(\mu's)$$
$$(11)$$

Note the important consistency requirement that our parameterization obeys: for x_0'=1., the free energy in eq. 10 is tantamount the free energy of an RHP with annealed crosslinks, a problem that has been recently analyzed [24]. Next, we compute numerically the stability of F with respect to x_0 and x_0' in all parameter regimes. While the value of ξ_0 obtained is used to determine the occurrence of few dominant folds, the stability of the free energy and the value of the parameterizations of ξ_0, x_0 and x_0', provides essential information on the sensitivity of the microdomain structure of the frozen phase to the quenched sequence distribution and the fixed crosslink realizations. Let us now explicitly explore the conformational organization of RHPs in the globular phase. The scenario depicted in fig. 1 corresponds to having a small number of cross-links of type A-B, and a positive χ_f (segments are encouraged to group with alike). The physics displayed by fig. 1 with regard to the conformational organization of the RHPs is qualitatively illustrated in fig. 2. At low T's, few dominant RHP folds occur due to formation of energetically driven microphases with segregated domain structure and having sharp interfaces. The heterogeneous cross-links formation, following the crosslinking procedure, nucleates at the A-B interfaces of the segregated microphases, and reduces the microphase interfacial free energy since each crosslink is an entropic constraint; thus chain reorganizations in this regime is most sensitive to the crosslink disorder component. The spatial microphase organization within the low temperature frozen phase is qualitatively represented in fig. 2; a heterogeneous cross-link is marked by two adjacent small

Fig. 2: Qualitative illustration of crosslink location among segregation patterns emerging within the frozen phase. Red and yellow regions depict microphases of A and B segments respectively, while the color intensity denotes the segregation strength. Two small adjacent circles represent two crosslinked monomers; red small circle represents a monomer of type A, while a yellow small circle represents a monomer of type B.

In the language of critical phenomena this would imply the occurrence of a coexistence regions between two frozen-globular RHP phases with an equal number of folds in each phase dominated in one phase by the sequence disorder and in the other phase by the cross-links disorder component. In the present system, ξ_0 is the only physically meaningful order parameter that measures the reduction in chain conformations, and thus monitors the occurrence of freezing; x_0 and x_0' parameterize ξ_0. As a result, the coexistence of separate individual such phases is not possible. Our numerical calculation confirms this fact, and the "coexistence" point does not occur; at T_1, a new first order conformational transition occurs to a more stable phase wherein the overall number of few dominant folds is abruptly reduced, as marked by the spike in ξ_0 (viz. red line in fig. 1); our numerical calculations show that this new globular phase is characterized by the occurrence of symmetry breaking induced by both crosslinks and sequence, thus we name it as mixed-frozen-globular phase. Let us now provide a physical interpretation of this conformational transition. Below T_1, the conformational organization of the RHP is primarily determined by the inter-segment interaction strength. In the vicinity of T_1 the existence of some cross-links induces formation of large domains bounded by cross-links. At the onset of formation of the mixed-frozen-globular state at T_1, the microphases are large in size and have diffuse interfaces (viz. fig. 2 - the seq. + cl. induced freezing). Since the crosslinks are heterogeneous, most likely they will form at this diffuse interfaces, decreasing the interface flexibility and entropy. But the penalty due to local deformation of large cross-linked bounded domains is small, and a statistical pattern matching of microphases belonging to separate cross-linked bounded domains occurs. The frozen globule occurs here due to both sequence and crosslink disorder components; only a few chain conformations allow prefered pattern matching of microphases over length scales as large as the size of cross-linked bounded domains, a realization which may explain the sharpness of the spike in fig. 1. The cooperativity observed here reminds to a large extent of the cooperativity observed in the folding of disulfide bonded proteins to their native state [15]. In the folding scenario of crosslinked proteins, crosslinks are also composition specific but homogeneous, they nucleate within the segregated domains. The apparently negative heat capacity inferred in fig. 2 is not inconsistent with thermodynamics; fig. 1 displays the temperature dependence of different thermodynamic systems and not phase behavior variation within the same thermodynamic system. This fact is a result of the experimental crosslinking procedure. At each temperature within equilibrated globular phases of linear RHPs, quenched crosslinks are formed. Thus at each T the system constraints (here the quenched crosslinks) are different which implies that the phase behavior comparison in fig. 1 is indeed between different thermodynamic systems at different temperatures. The complementing future study scenario, is the phase behavior of a linear RHP system crosslinked at one temper-

ature, and subject to temperature variations while the crosslink realization of the initial crosslinking temperature is retained.

Let us now return to the phase behavior analysis of fig. 1. At T_2 (viz. fig. 1), our numerical calculation shows that the mixed-frozen-globule becomes unstable, while the sequence-frozen-globular phase is stable and it should be observed as it has the lowest chemical potential (viz. fig. 1 green line). The occurrence of this continuous conformational transition is expected; the number of cross-links is small while the reduction in the number of dominant folds due to cross-links only cannot occur at high temperatures. The $T > T_2$ frozen regime should be characterized by weaker and diffuse microphases (viz. fig. 2 - sequence induced freezing). A good candidate for testing our predictions is the ensemble growth Monte Carlo method [13]. This approach has been successfully implemented in the study of linear RHPs in charged disorder [25], [26] and in confined geometries, and allowed a faithful comparison with prior analytical calculations on RHPs [2], [6].

Acknowledgments

[1] R. Deam and S. F. Edwards, Pro. Tans. R. Soc. London Ser. A 280 (1976) 317

[2] S. Panyukov, Sov. Phys. JETP 69 (1989) 342

[3] P. M. Goldbardt, H.E. Castillo and A. Zippelius, Adv. Chem. Phys. 45 (1996) 393

[4] E. I. Shakhnovich and A. M. Gutin, J. Phys. (Fr) 50 (1989) 1843

[5] G. H. Fredrickson and S. T. Milner, Phys. Rev. Lett. 67 (1991) 835

[6] C. D. Sfatos, A. M. Gutin and E. I. Shakhnovich, Phys. Rev. E 48 (1993) 465

[7] V. S. Pande, A. Y. Grosberg and T. Tanaka, Phys. Rev. E 58 (1995) 4

[8] S. Ramanathan and E. Shakhnovich, Phys. Rev. E 50 (1994) 1303

[9] E. I. Shakhnovich and A. M. Gutin, Biophys. Chem. 34 (1989) 187

[10] V. I. Abkevich, E. I. Shakhnovich, Biochemistry 33 (1994) 10026

[11] A. Irback and H. Schwarze, J. Phys. Math. Gen. 28 (1995) 2121

[12] H. S. Chan and K. A. Dill, J. Chem. Phys. 12 (1994) 100

[13] P.G. Higgs and H. Orland, J. Chem. Phys. 95 (1991) 4506

[14] K. Binder and A. P. Young, Rev. Mod. Phys. 58 (1986) 801

[15] T. E. Creighton, Protein Folding (W. N. Freeman and Company New York 1992)

[16] S. F. Edwards, Proc. Phys. Soc. London 85 (1965) 613; K. Freed, Renormalization Group Theory of Macromolecules (J. Wiley & Sons, 1987)

[17] L. Gutman and A. K. Chakraborty, J. Chem. Phys. 101 (1994) 10074; 103 (1995) 10733

[18] J. F. Joanny, J. de Physique II 4 (1994) 1281

[19] P. Goldbart and N. Goldenfeld, Phys. Rev. Lett. 58 (1989) 2676

[20] P. Goldbart and N. Goldenfeld, Phys. Rev. E 39 (1989) 1402; 39 (1989) 1412

[21] A. M. Gutin and E. I. Shakhnovich, J. Chem. Phys. 100 (1994) 5290

[22] L. Gutman and E. I Shakhnovich, J. Chem. Phys. (accepted)

[23] G. Parisi, J. Phys. A: Math. Gen. 13 (1990) 1887; M. Mezard and G. Parisi, J. Phys. I 1 (1991) 809

[24] L. Gutman and E. I. Shakhnovich, J. Chem. Phys. 107 (1997) 0000

[25] D. Bratko, A. K. Chakraborty and E. I. Shakhnovich, Phys. Rev. Lett 76 (1996) 1844

[26] S. Srebnick, A. K. Chakraborty and E. I. Shakhnovich, Phys. Rev. Lett. 77 (1996) 3157